

Reprinted from IEEE TRANSACTIONS
ON ELECTRONIC COMPUTERS
Volume EC-15, Number 1, February, 1966
Pp. 108-111

Copyright 1966, and reprinted by permission of the copyright owner
PRINTED IN THE U.S.A.

Time Loss Through Gating of Asynchronous Logic Signal Pulses

IVOR CATT

Abstract—The gating of asynchronous signals causes logical errors. It is possible to reduce the frequency of these errors, but the price paid is a severe loss of time and extra cost in hardware.

Consider Fig. 1. Suppose B is an asynchronous signal entering some clocked logic with clock A , and it is brought into synchronization by ANDing ($A \cdot B$). There is a statistical possibility that chaos will result, as indicated by the special case in Fig. 1. For example, line G might indicate that data transfer had taken place at time t_1 ,

although line H only enabled the transfer to take place at time t_2 .

A first (insufficient) step is to make C drive a bistable, as in Fig. 2. However, there is still a (smaller) statistical possibility that chaos will result, as indicated by the special case in Fig. 2, because the flip-flop may enter its metastable (half-set) state for an extended period of time. As a result, the system will still get out of step. For discussion of this metastable state, see Appendix I.

The gating of asynchronous signals will always carry a statistical chance of logical failure, because it carries with it the risk of the appearance of a half-amplitude, half-true logic signal. The inter-

Manuscript received February 18, 1965; revised August 30, 1965, and October 13, 1965.
The author is with the Semiconductor Products Division, Motorola Inc., Phoenix, Ariz.

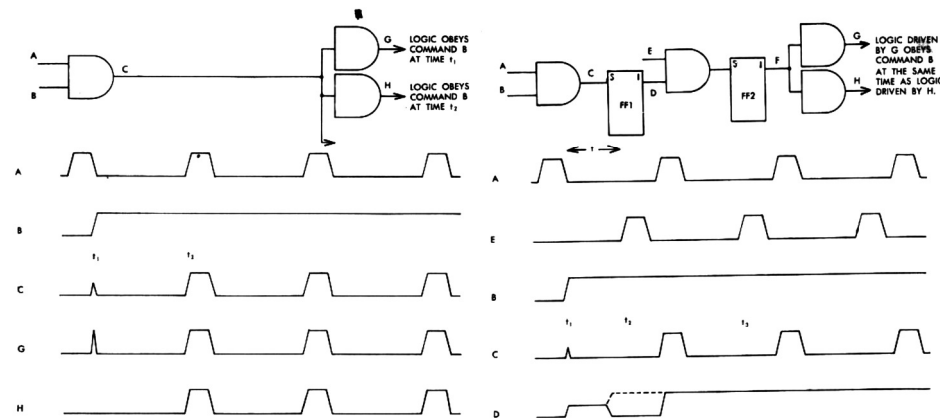


Fig. 1.

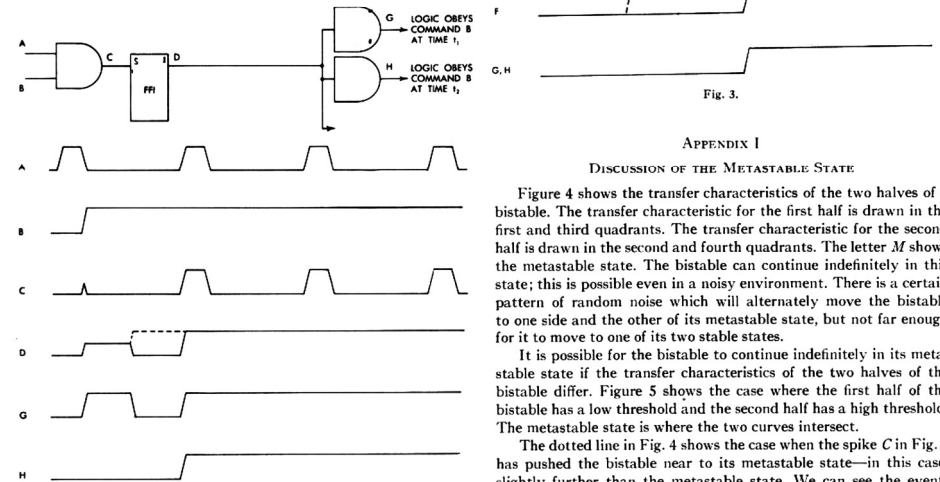


Fig. 2.

mittent failure resulting is more serious than a catastrophic failure, because it introduces undetected errors.

The following is a description of how to reduce the failure rate to acceptable proportions. A good rule of thumb for acceptable failure rate for a logical circuit is one failure in about 1000 days, or three years.

Figure 3 shows the circuit. ($A \cdot B$) tries to set FF1 at time t_1 . At time t_2 , after a delay of length t_d , the output of FF1 is interrogated, and it is hoped that FF1 will by that time have left its metastable state and returned to one or other of its stable states, so that during one of the clocks E , FF2 will be set unambiguously. The probability that at time t_2 , FF1 will still be metastable and, therefore, that a failure will occur in spite of its protective circuitry, is computed in Appendix II. It is easy to make this probability acceptably low. Unfortunately, the time ($t_2 - t_1$) lost in the process is uncomfortably large.

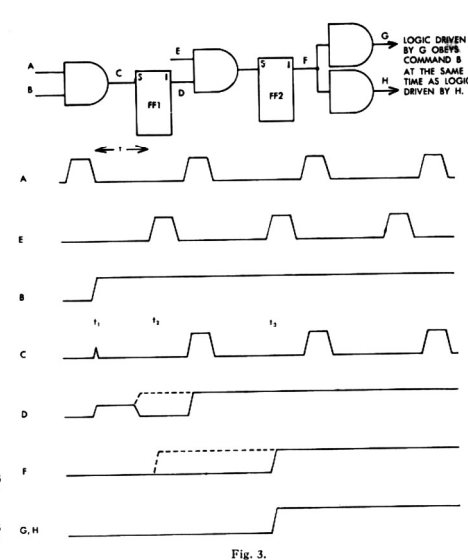


Fig. 3.

APPENDIX I

DISCUSSION OF THE METASTABLE STATE

Figure 4 shows the transfer characteristics of the two halves of a bistable. The transfer characteristic for the first half is drawn in the first and third quadrants. The transfer characteristic for the second half is drawn in the second and fourth quadrants. The letter M shows the metastable state. The bistable can continue indefinitely in this state; this is possible even in a noisy environment. There is a certain pattern of random noise which will alternately move the bistable to one side and the other of its metastable state, but not far enough for it to move to one of its two stable states.

It is possible for the bistable to continue indefinitely in its metastable state if the transfer characteristics of the two halves of the bistable differ. Figure 5 shows the case where the first half of the bistable has a low threshold and the second half has a high threshold. The metastable state is where the two curves intersect.

The dotted line in Fig. 4 shows the case when the spike C in Fig. 3 has pushed the bistable near to its metastable state—in this case, slightly further than the metastable state. We can see the events that follow by going clockwise, starting with the first quadrant. In this case the bistable gets out of the transition region in a time $2t_d$, or twice round the loop in the bistable.

The practical case of a metastable state will be more complex than the steady-state picture described previously. In practice, the metastable state might be represented by an oscillation of the bistable outputs about a mean level represented by the steady metastable state. This is the situation investigated in Appendix II.

APPENDIX II

CALCULATION OF FAILURE RATE

In Fig. 3, let us assume that the number of times per second that B goes true (and, therefore, there is a chance of failure) is f . A reasonable figure for f when multiplexing data transfers from a number of tape stations into a computer would be 10⁶.

If we are considering a three-year period (equals about 10⁸ seconds), then the number of times B goes true is about 10¹⁴. If

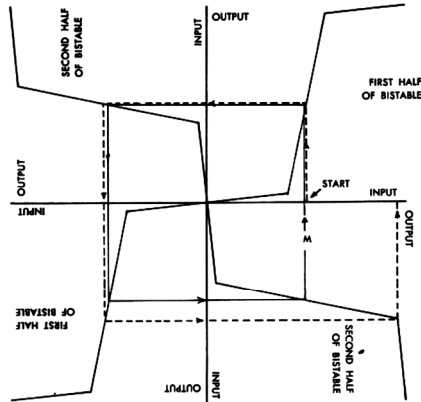


Fig. 4.

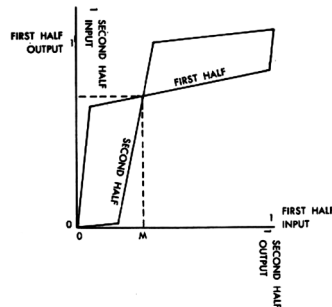


Fig. 5.

we allow one failure during this time, we allow a failure probability of $1/10^6 \times f$.

The first bistable is driven into its metastable state if the width of the spike occurring at C at time t_1 is less than the delay around the loop in bistable FF1. If the transistor frequency bandwidth is F , then the delay around the bistable will be approximately $10/F$ seconds.

If B goes positive f times per second and $f \ll F$, then the proportion of times that the trailing edge of A catches B and generates a spike is

$$\left[\frac{\frac{10}{F}}{\frac{1}{fA}} \right] = \frac{10f}{FA}$$

Thus, the number of times in three years when a small spike occurs is

$$10^8 f \times \frac{10/A}{F} = \frac{10^9 f A}{F}$$

Let us assume that

$$R = \frac{\text{Transition region amplitude}}{\text{Logic swing}}$$

Then R represents the proportion of the small spikes which will be of such amplitude as to partly succeed in setting bistable FF1, and so FF1 enters its metastable state $(10^8 f^2 \times R)/F$ times in three years.

We now have the picture of a roughly half-sized pulse recirculating around the feedback loop of the bistable FF1. We may assume that there is a regular distribution of amplitudes among these pulses, starting with small ones which just get into the transition region, through pulses nearly half way through the transition region, to large pulses which are only just below the top of the transition region.

Each time this pulse passes around the loop, it is amplified twice. If the amplification in each half of the bistable is β , the amplification around the loop is β^2 . In one trip round the loop, a number of pulses will drop out of the transition region. After n trips round the loop, the number of pulses still within the transition region will decrease by a factor $(1/\beta)^{2n}$.

Now we must wait so long that only one pulse remains in the transition region during three years. This means that

$$1 = \frac{10^8 f^2 R}{F} \times \left[\frac{1}{\beta} \right]^{2n}$$

Therefore, the number of times that the pulse must be allowed to circulate round the bistable loop is

$$n = \frac{1}{2} \frac{\log \left[\frac{10^8 f^2 R}{F} \right]}{\log \beta}$$

The time t in Fig. 3 is $n \times d$ where d is the delay round the loop of the bistable.

Example

As an example, let us substitute the following values into the formula:

Number of times per second that signal B arrives at $f = 10^6$

$$\frac{\text{Transition width}}{\text{Signal logic swing}} = R = \frac{1}{10}$$

Frequency bandwidth of the transistors $= F = 10^9$

Gain of one half of the bistable $= \beta = 5$

Delay round the bistable $= d = 10/F = 10$ ns. Then,

$$n = \frac{1}{2} \frac{\log \left[\frac{10^8 \times 10^{10} \times \frac{1}{10}}{10^9} \right]}{\log 5} = \frac{1}{2} \times \frac{20.7}{1.61} = 6.4$$

Now $d = 10$ ns,

$\therefore t = n \times d = 64$ ns.

Worst-case delay

The worst-case (longest) delay will be greater than 64 ns for two reasons:

1) In Fig. 3, the worst-case delay is $(t_2 - t_1)$. This equals $2t$ plus (say) 20 ns. This 20 ns allows for the length of pulses E and A and also the gap between them, so thus, we have a delay of $2t + 20$.

2) The 64 ns was computed on the basis of various tenuous assumptions. The author considers that for safety, it should be doubled. This gives a total delay of $4t + 20 = 276$ ns.

Thus, we end up with a worst-case delay of about 300 ns. Note that this is when using transistors with a frequency bandwidth of the order of 1 Gc/s.

APPENDIX III

DISCUSSION OF THE SOLUTION TO THE PROBLEM

Figure 3 shows the circuit and timing diagram which solves the problem as far as it is possible to solve it. What is necessary is to get the logic signals out of the transition region as they enter the system. The basic element of a successful circuit is a very high-gain digital stage. That is, we need an element with a very narrow transition region. This can be achieved by placing a number of logic elements in series. However, it seems easier to use only two logic elements cross coupled. The result is a bistable FF1.

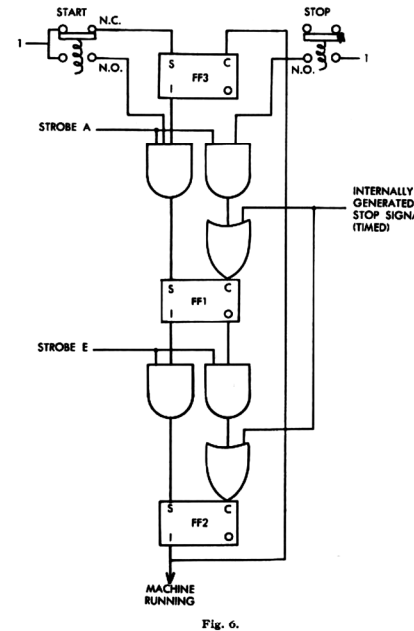


Fig. 6.

APPENDIX IV

LOGIC AND CIRCUITRY TO BACK UP THE START PUSH BUTTON AND THE STOP PUSH BUTTON

Figure 6 shows an example of circuitry used to back up switches and push buttons in a digital system. The timing of Strobe A and Strobe E is as in Fig. 3. Bistables FF1 and FF2 perform the same function as they do in Fig. 3. Bistable FF3 insures that if an internal STOP signal should be generated very soon after the manual depression of the START signal, the machine will not start a second time.

LETTERS TO THE EDITOR

BETTER R.F.I. PROTECTION NEEDED

It is clear from my own observations that a.m. citizens' band equipment operating on the 27MHz frequency is now so firmly entrenched in this country that nothing, certainly not the belated appearance of a legal specification, will sweep it away. Whatever the rights and wrongs of the matter may be, there are just too many a.m. rigs in service for them to fade rapidly into obscurity come the glorious day.

I therefore issue a vehement plea for all manufacturers of domestic electronic equipment to start looking seriously at one aspect of its performance which is usually wholly neglected — immunity to strong radio-frequency fields. Manufacturers ought to be forcefully reminded that if their apparatus is not intended to respond to 27MHz a.m. signals it is a failing on their part if it does. The extra components needed to secure excellent r.f.i. protection are not expensive, and their presence would also assist in reducing the number of domestic problems arising from the use of amateur, p.m.r., broadcast or other radio transmitters close to ordinary households.

Perhaps reviewers might observe that an r.f.i. susceptibility test would be a useful addition to their array of measurements. A number of reputable hi-fi manufacturers produce amplifiers with appalling r.f.i. protection, and it seems that performance in this respect is haphazard — there being considerable differences between various models from the same manufacturer, and no apparent correlation between price and protection.

Norman McLeod
Brighton
Sussex

DISTORTION AT THE AMPLIFIER-SPEAKER INTERFACE

The two-part article "Intermodulation distortion at the amplifier-loudspeaker interface" by Otala and Lammasniemi in your November and December 1980 issues contains serious flaws.

This article began life as an Audio Engineering Society Convention preprint, No. 1336 of February/March 1978. Its authors are aware of at least three independent rebuttals of that preprint, one of which has already been published. This published rebuttal is by R. R. Cordell of Bell Telephone Laboratories, and is available as AES Convention preprint No. 1537 of November 1979, under the title "Open-loop output impedance and interface intermodulation distortion in audio power amplifiers". One of the unpublished rebuttals is by E. M. Cherry and G. K. Cambrell of Monash University; originally submitted to the AES Journal in February 1979, a revised manuscript was submitted in October 1980 under the title "Output stages for audio power amplifiers".

Cherry and Cambrell make the following points:

1. If an amplifier uses a common-emitter output stage then, if collector resistance can be varied without changing any other parameter, interface intermodulation distortion, i.e., increases

monotonically as collector resistance is reduced.

2. If an amplifier using a given transistor has a common-emitter output stage, and if this is changed to the common-collector configuration and nothing else is changed except the phase of the feedback connection, i.e., at best remains constant but is more likely to increase.

Taken together, 1 and 2 run absolutely counter to the suggested "rule" of providing a low open-loop output resistance (WW Dec. 1980, p. 56).

3. For practical purposes, a loudspeaker is passive and cannot inject a signal back into an amplifier. (a) The motional e.m.f. produced by incident on the loudspeaker cone from room or enclosure reflections of from other sources is minuscule compared with amplifier rated output voltage. (b) Substantial motional e.m.f. results from the signal applied to a loudspeaker. However the substitution (or compensation) theorem of network theory shows that an active network which models a loudspeaker and includes such a motional e.m.f. can be replaced identically by the passive LRC network that completely models the driving-point impedance of the loudspeaker. A loudspeaker is strictly passive so far as any applied electrical signal is concerned, and there is no possibility of i.i.m. as defined because there is no independent signal source in the load.

4. I.i.m. is proportional to a product of output current amplitudes in Fig. 4. The constant of proportionality depends on the detail of the circuit, but cannot exceed the constant in a given two-tone intermodulation test. I.i.m. at standard output current amplitudes cannot exceed standard intermodulation at the same current amplitudes.

Taken together, 3(a) and 4 suggest that the distortion power produced in a real-life situation by the interface intermodulation mechanism is minuscule compared with the distortion power produced by the standard intermodulation mechanism.

Edward M. Cherry
Department of Electrical Engineering
Monash University
Clayton, Victoria, Australia

The authors reply:

We are not aware of any rebuttals of our AES paper. The paper of Cordell is based on different premises from ours, i.e., Cordell postulates the amplifier open-loop distortion to be constant in the comparison, whereas our analysis is based on the closed-loop distortion being held constant. This difference in boundary conditions taken into account, Cordell's results are in agreement with ours and the paper can hardly be considered a rebuttal. The two other references quoted are unknown to us, and will be considered if and when available.

The points the writer makes sound familiar to us as if they were our own results taken from our paper:

1. This conclusion is a corollary to our paper. We assumed the amplifier closed-loop distortion to be constant, which is a real-life engineering consideration, as discussed in our paper. The writer's assumption is that the open-loop distortion is constant and that the amount of overall negative feedback varies with the collector resistance. This leads to complete agreement with

our results, if allowance is made for the different boundary conditions. However, we doubt if the writer's case could be realistic in practice.

2. Our theory shows that the i.i.m. in this case should in principle remain about the same just as the writer states. We cannot see any theoretical discrepancy here either. Nevertheless, this kind of a hat-trick would be impossible in practice, and practical measurements show the common-emitter stage to be inferior because of larger closed-loop distortion.

3. (a) We agree completely with this point, as is stated in our paper. (b) As far as the loudspeaker is concerned, this is just a matter of definition. We would wish to point out that the proposed i.i.m. measurement method was not conceived to simulate the physical loudspeaker, but just to expose the amplifier output port to such worst-case current and voltage relationships which might occur when real loudspeaker loads are being driven.

4. This is a rephrasing of the opening paragraph of Part 2 of our paper. In many cases, i.i.m. will be negligible as compared to the CCIF two-tone i.i.m. However, in a poorly designed amplifier, such as shown in our Fig. 14, it may equal in magnitude the two-tone i.i.m., as can be seen from our Figs. 15 and 17.

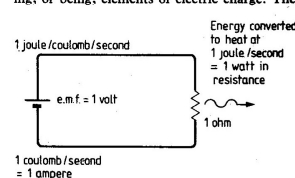
In conclusion, the letter does not seem to indicate any flaws in our paper, on the contrary. Many a thing may seem controversial if viewed from different positions. However, a more thorough examination which takes into account the different sets of boundary conditions shows no conflict to exist.

Matti Otala, Torma Lammasniemi
Technical Research Centre of Finland
Oulu, Finland

THE DEATH OF ELECTRIC CURRENT

Mr Ivor Catt's very interesting article in your December 1980 issue obviously calls for some discussion, since, if he is correct in his analysis it would imply that a lot of our fundamental teaching in electronics is wrong.

Let me recapitulate first, simply, on the Normal theory of electric current flow. It is now widely taught that in the following circuit the electric current consists of a flow of electrons, between adjacent atoms which make up the material of the wires; the electrons either carrying, or being, elements of electric charge. The



charges are given energy by the electromotive force of the battery, such that if 1 coulomb (6.24×10^{18} electrons) of charge is raised through a potential difference of 1 volt, it acquires 1 joule of energy; which is then expended when the current (rate of flow of charge) flows through the external circuit resistance. If the charge is

flowing through the wire at 1 coulomb/s, then the current is said to be 1 ampere, and the resistance of the circuit would be 1 ohm; while the energy of the current would be dissipated (e.g. converted into heat) by resistance, at the rate of 1 watt, or 1 joule/s.

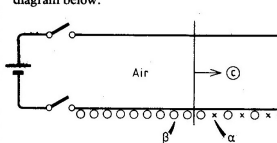
It would seem from the successes we have had, for example, in making colour television, radio and stereo systems available to so many people, that these circuit fundamentals must be quite a valid and useful way of thinking. I am also at a loss to see how Mr Catt can develop his theory of the battery and resistor, with the 'energy current' entering the resistor sideways (on p. 80, December issue) into giving such useful quantitative concepts as the above circuit does; but maybe he doesn't want to, at present. It would seem, however, that he is at least asking us to lay aside our hypotheses about the existence of protons, electrons, and therefore presumably even atoms; for we are told that electric charge does not exist, and nothing flows in a conductor. This could indeed be revolutionary.

As a philosopher, I am only in sympathy with Mr Catt's initiative. Although I can't really follow the flight of his imagination at present, I have argued elsewhere ("Mind & Machine," *The Listener*, Oct. 17th, 1963) that the concepts and inventions of physics, and indeed the Universe itself, should be understood in terms of the concept of imagination, e.g. of the writing of scientists, and not vice versa. My attempt to argue this viewpoint however, i.e. that scientific knowledge does not have to be taken literally as ultimate truth, was not very well received, and I was accused of 'dangerous obscurantism'. It may, I suppose, one day be possible to explain the 'imaging' or 'imagining' function of the brain in physical concepts. However, although I wish Mr Catt every success in developing his imagination and new theories, I think he should be warned, or reminded, that the imagination of scientists does have to be supported, or tested, by observations and experiments. In short, it seems that he may be unwise in reviving a Heaviside theory, published in 1892, and in quoting J. A. Fleming (1898) and Clerk Maxwell (1831-1879), who lived before the discovery of the electron (1897), through the experiments of J. J. Thomson, had become well known and accepted.

Peter G. M. Dawe
Oxford

The author replies:

Mr Dawe's recapitulation, para. 2, deals with a so-called 'steady state' situation. Conventional theory covers for these quite well; it was developed for that purpose. However, conventional theory cannot cope with the transient condition, as we shall see. Consider the situation $\frac{1}{4}$ nanosecond after we close the switches in the diagram below.



A voltage-current step has advanced three inches to the right. Behind the step, there is a voltage drop between the wires. The E lines must terminate on electrons in the lower wire. It follows that behind the step the lower conductor contains more electronics per inch than is contained in the uncharged section ahead of the step.

As the step advances further forward, extra electrons must appear in locations such as α to terminate the new E lines involved in the voltage difference which now exists in the next inch of transmission line.

Where does the electron come from to fill the next gap α as the step front advances forward? It cannot be one (say β) from behind the step, because this electron is not travelling at the speed of light. For β to arrive at location α in time, it would have to travel at the speed of light, since the voltage-current step is travelling forward at the speed of light (for the dielectric). A central feature of conventional theory (N or H) is that the drift velocity of electric current is slower than the speed of light. Therefore Theory N, where electric current is the cause and $E \times H$ field an effect, breaks down for the simple reason that a cause travelling slower than the speed of light cannot create an effect travelling at the speed of light. It seems clear that if we retain a dualistic theory (N or H), the present discussion forces us to conclude that Theory H obtains; the cause must be the $E \times H$ field in the dielectric, energy current, which does travel at the speed of light, and the slower electric current in the wire is merely an effect of that cause.

I would agree with Mr Dawe, para. 3, that practical success would tend to indicate that our fundamental theory is sound. However, counter-examples abound. Lacking sound theory, the Romans still built many impressive bridges. Like Mr Dawe, I shall use whatever suits me to calculate dissipation in resistors, etc. We do not have to use the theory we believe, when it is inconvenient, rather than travel by another more convenient path in our day-to-day affairs. Calculation of the steady current from a (car) battery to a resistor (car headlamp) will not become the stamping ground for theoretical discord. Similarly, I think quite happily about how to avoid "losing the cold" in my deep freeze. There is a time and place for theories. The policeman who charges you with driving without due care and attention should not have to bother with Newton's Laws of Motion, and is not charging you for ignoring them.

With regard to the last paragraph, the electron is not necessary (indeed, it creates major problems) in explaining the passage of a TEM step guided between two conductors. Should it be necessary in other situations, it can be expected to turn out to be a standing wave energy current. This was proposed by Schrödinger. Jennison's design of such a structure (*Wireless World* June 1979, pages 45-47) goes wrong because, like so many others, he is trapped within the conceptual confines of the sine wave. Once you drop the sine wave, it is not difficult to construct an "electron" out of energy current. (However, it would then be illogical to hold onto Theory N or Theory H, since energy current would then be bordered by energy current (i.e. electrons). Similarly, once it is realized that a capacitor is a transmission line, it is not logical to retain the alternate lumped L and C (transmission line) model for the transmission line.)

I think the first part of the last paragraph, like Osiander, is wrong. It is a tragedy that virtually all contemporary scientists are siding with the mediaeval church against Galileo. I stand with Galileo, Bruno and Kepler, but unlike Bruno I shall not be burnt alive for it. (See M. Polyanyi, "Personal Knowledge", RKP 1958, pp. 145-6.) As to the second part of the last para., I am making *discovery*, not indulging in imagination. As to the electron, although I may allow the existence of the standing-wave electron, I find the billiard-ball electron incomprehensible. Like Einstein, I do not accept the quantum. (Max Born, "The Born-Einstein Letters", Mac.

millan 1971, pp. 164, 168.) However, this does not bear directly on Theory C, which merely removes the (possibly in other situations surviving) electron from the theories of (a) the "steady charge capacitor" and (b) "electric current in a wire".

Ivor Catt

HERBERT DINGLE

Perhaps I may be permitted to make a brief reply to Dr Wilkie's lengthy attack in the June issue on my late uncle Professor Herbert Dingle. Dr Wilkie writes: "Professor Dingle is described as an expert on relativity". He makes no comment on this but later in his letter he says "Professor Dingle was a distinguished historian of science". The subtle implication is that he must be regarded as an historian who had no right to be delving into such abstruse matters as the Theory of Relativity. This impression can best be corrected by quoting from his obituary in *The Times* of September 6th, 1978.

"His 'Relativity for All' (1922) appeared at a time when it used to be said that only six men in the world understood the theory. If this had been true, Dingle must be rated high among the six for his little book showed a profound grasp of relativity as a physical theory combined with a capacity for presenting it, not as an esoteric mystery, but as a logical development of the mechanics of Newton".

To this might have been added the comment that he met and discussed scientific matters with Einstein, a privilege that was denied to most of his critics.

My other point concerns my uncle's love of good English. This was something he inherited from his father and shared with his brother. It led him to avoid jargon whenever possible. Dr Wilkie, who evidently loves technical language, finds this very tiresome; he holds the remarkable view that plain English is ambiguous and jargon is precise. I know from my own profession as a veterinary surgeon just how mistaken this is. Once people resort to jargon they make words mean whatever they want them to mean; one only has to recall what happened to 'parameters' to realise that.

I have not the knowledge to tell whether my uncle's beliefs were correct, but I confess I am not impressed by an opponent who admits to difficulty in expressing his case in plain English, and who links Herbert Dingle's supporters with people who believe the Earth to be flat. 'Flat Earthers', by the way, can be dealt with quite easily without resorting to technical language.

P. J. Dingle
King's Lynn
Norfolk

TELEVISION SETS FOR THE DEAF

I am glad that Mr Power has pointed out that hearing impaired people will not necessarily get satisfactory listening via a manufacturer's installed outlet socket (May letters). When 15 per cent of the adult population have hearing difficulties it seems appalling to me that none of the manufacturers pays attention to the problem.

I wrote my original letter to you with my tongue just a little in my cheek as I know more than a little about the problem. I was hoping to draw a hail of fire from the various manufacturers but only Desca had anything to say.

May I conclude by saying that the problem is not for the hearing impaired alone; it is a problem for their families and neighbours as well. One of the most common enquiries which I get

THE SCIENTIFIC REFEREE SYSTEM

M.H. and B.R. MacROBERTS

740 Columbia, Shreveport, Louisiana 71104, USA.

Received: 16 January 1980

Abstract

There has been very little written about the scientific referee system but a lot has been implied. It seems to be widely believed that the system works well, even though there are cases of disparate judgement. These however are usually explained away in an ad hoc fashion. We find that novelty is characteristically resisted by scientists and suggest reasons for this.

1. INTRODUCTION

The referee system began in the seventeenth century in England in the form of a collator who saw papers through the press. Because of the low quality of many manuscripts, this system was soon replaced by editing and refereeing. The practice of evaluating the substance of manuscripts soon developed. Referees were chosen on the basis of their expertise. "Almost from their beginning, then, the scientific journals were developing modes of refereeing for the express purpose of controlling the quality of what they put into print."⁽¹⁾ This is the current practice. Each journal has an editor who receives and judges papers or passes them on to reviewers for judgement.

Because short articles in journals have become the major publication method in science, referees and, to a lesser extent, editors have become the "lynchpin" about which the system pivots. This being so, it is not difficult to see the importance of editors and referees to the process of scientific communication and scientific advancement. The question that we raise here — and one raised by others — is how well does the system work?

2. THE RECEIVED VIEW

The majority opinion, which we will call the "received view", is that although the referee system does not work unfaithfully, it approximates the ideal. This view is implicit or explicit in much of the writing of those sociologists who make the institution of science their speciality.^(2,3) The assumption underlying this view is that scientists are largely aparaadigmatic thinkers, testing ideas against nature. The individuals who act as referees are believed to be objective, disinterested, sceptical, sympathetic, open, tentative and hospitable to change.⁽⁴⁾ Because of this and the fact that they are "experts", referees should be nearly infallible judges of work that falls within their areas of specialization. Papers published in "reputable" journals therefore

"bear the *imprimatur* of scientific authenticity" because they do not merely represent the opinions of their authors but also those of the editor and the referees.⁽⁵⁾ Consequently, the practice "of monitoring scientific work before it enters into the archives of science means that much of the time scientists can build on the work of others with a degree of warranted confidence"⁽⁶⁾.

According to the received view, papers submitted for publication fall along a simple continuum. At one end are the very bad ones, written presumably by dullards; at the other are a few excellent papers, the "cognitive products" of the elite. In between are the majority, "hod work", as Darwin⁽⁷⁾ called it. The editors' and referees' task is simply to assign each paper to its position and decide on a cut-off point.

Adherents to the received view, however, do not overlook the fact that the system does not work with unfaithful effectiveness, but — and this is the point — their treatment tends to dismiss incompetent judgements as *ad hoc* events. Consequently, the received view does not address itself to the possibility that intellectual bias is an important factor in the assessment and acceptance of scientific work.

Traditionally and typically, supporters of the received view have painted a reassuring picture, one mirroring the Panglossian mode in which almost everything that happens, recurrent or otherwise, is seen as basically for the good. Anomalies are distorted into isolated and unusual events or are sometimes even construed as being of positive value, that is, when anomalies are admitted, some institutional necessity is used to justify them.⁽⁸⁾

Therefore, it would seem that the received view would predict that the rejection of important, novel work should be infrequent. Certainly important work should not be repeatedly rejected. Figure 1 (below) approximates what we take to be the received view. Scientific progress in this view is easily understood.

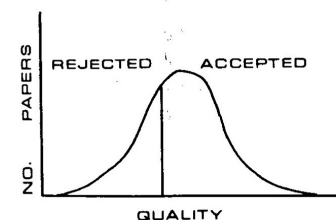


Figure 1. The Received View.

3. THE MULTI-PARADIGM MODEL

The necessity for a different interpretation than is provided by the received view begins with the realization that the frequency with which good papers are rejected and the extent to which scientists resist new ideas are not accounted for by the received view.⁽⁹⁾ Inconsistencies between what occurs and what is predicted to occur are too major and too frequent to be explained away as occasional mistakes. In fact, what the "mistakes" suggest is that they are not mistakes at all, for as Taton⁽¹⁰⁾ has shown, "the number of revolutionary discoveries which came into their own, only after hard battles, is legion."

We therefore reject the received view and propose another interpretation of these events.⁽¹¹⁾ According to this interpretation, scientists are not unbiased, objective, sceptical, disinterested, sympathetic, open, tentative or hospitable to change, but the opposite. Scientific judgement is in terms of prevailing opinion. It is paradigmatic.⁽¹²⁾ Put differently, scientists are "encapsulated"⁽¹³⁾, or as Hoyle⁽¹⁴⁾ has said, "straight jacketed". Kantor⁽¹⁵⁾ puts it simply, "established doctrine in science is more powerful than factual evidence".

Although Kuhn is responsible for popularizing the notion of paradigmatic science, this conceptualization has been around for a long time and has been variously expressed, for example by Kantor⁽¹⁶⁾ and by Koestler⁽¹⁷⁾. Koestler⁽¹⁸⁾ perhaps has captured the idea behind the concept of paradigmatic commitment as well as any. He sees typical paradigmatic behaviour as involving a "cognitive matrix with a distorted logic, the distortion being caused by some central axiom, postulate or dogma, to which the subject is . . . committed, and from which the rules of processing the data are derived". He⁽¹⁹⁾ continues, "The amount of distortion involved in the processing is a matter of degrees . . . it ranges from the scientist's involuntary inclination to juggle with data as a mild form of self-deception, motivated by his commitment to a theory, to the delusional belief-systems of clinical paranoia . . . But," he concludes, "to undo a mental habit sanctified by dogma or tradition, one has to overcome immensely powerful intellectual and emotional obstacles."⁽²⁰⁾ Thus the scientist is committed to a world view from which reality is constructed, and things are fitted or "twisted" as Kantor⁽²¹⁾ has said into conforming with that view.

Scientific work in these terms can be depicted as a three-dimensional field, as in Figure 2 (below). The large hill in the foreground represents the major paradigm of a discipline, for example, Functionalism in sociology. Most scientists accept the major paradigm, and when judging the work of others, they do so on the basis of its approximation to the standards of the paradigm. To the rear of the field are several small hills, which symbolize small coterie pioneering new approaches. Members of such coterie are working under assumptions different from those of the majority. These assumptions are not understood by adherents to the major paradigm, yet members of the coterie understand the major paradigm, for they work out from it — in fact, it is usually this they are rejecting.

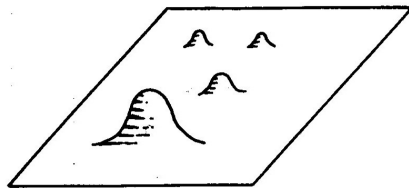


Figure 2. Multi-paradigm model

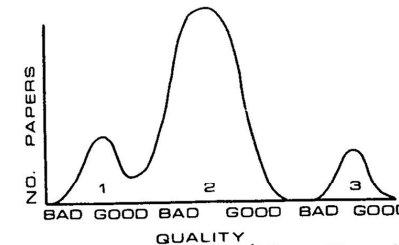


Figure 3. Two-dimensional view of the multi-paradigm model.

Visualized two-dimensionally, scientific papers do not fall into a single bell-shaped curve but rather into a series of curves in which there are good and bad papers (see Figure 3 above). Most referees can only judge papers that fall into one or two curves. But they attempt to judge all papers on the basis of their own standards. For example, if the referee is a mentalistic psychologist, to suggest that the current view of the nature and function of the central nervous system is traceable to the Patristics and is completely ascientific is to kill the paper before it is even born. Again, to suggest that Newton was badly confused about the part played by light in vision and that instead of seeing light, we see things by the medium of light, is to ask for trouble, and so on.⁽²²⁾

What is being said by a heretic seems ridiculous to adherents of the accepted paradigm. To them, papers that are novel often appear to be unpolished, simple-minded, improperly referenced, misguided, poorly conceived, amateurish, or just plain wrong, depending on the degree to which they deviate from the paradigm. Their authors are considered, as Van Valen and Pitelka⁽²³⁾ put it, to be "presumptuous". But the reason for this is that the new knowledge or point of view is not part of the paradigm or goes directly counter to what is thought to be "fact". Therefore, the innovator is judged wrong or incompetent by the fact of showing glaring deviations from what is accepted or acceptable.

Storer⁽²⁴⁾ captures the essence of the problem when he says, "A contribution must be an understandable extension of knowledge, and if it fails to meet this test, either because others simply cannot understand its relation to current knowledge or because it seems to fly in the face of accepted standards of reasoning, it will be rejected and its author's credentials as a scientist will become suspect. This means that a scientist who is so far ahead of his colleagues that they cannot understand how he arrived at his conclusions or why they are important will be treated in the same way that a hollow-earth theorist is treated."

Because journals are usually controlled by adherents to the major paradigm, the consequences of submitting non-paradigmatic work are predictable.

4. CONCLUSION

While the referee system came into being in order to separate the trivial and incompetent from the competent at a time when it may have been possible and perhaps necessary to do so, the system also immediately became an

obstacle to creativity and innovation, for although it does sort, it does so not only on a good-bad basis but also on a paradigmatic-nonparadigmatic basis, and herein lies its great and inherent weakness.

Whatever positive aims were envisaged by instituting the referee system and whatever positive goals it achieves, the seeds of dysfunction were sown at its inception, for given the paradigmatic nature of science and the encapsulation of scientists, the natural outcome can only be resistance to innovation and rejection of novelty. The weak links in the system are the referees themselves, for they like all scientists have repeatedly shown themselves to be unable "at any given time to distinguish between an idea that is entirely wrong and an idea that may be received as brilliant at some later date"²⁵).

In conclusion, let it be said that even though it would be hard to find a scientist who is not aware that some good papers are rejected and some novelty is resisted, it would also be hard to find a scientist who altered his own behaviour because of this knowledge. It is not surprising that individuals who have not experienced resistance accept the received view. Not having had work rejected for paradigmatic reasons, it is easy for them to say that this either does not occur or that it represents an unusual event in science. The Nageli-Mendel incident illustrates the problem perfectly. Here is the "Father of Genetics" trying patiently to explain what he has found to the expert on plant hybridization, only to be rebuffed and sent back to his garden. If supporters of the received view can only imagine this incident multiplied a thousand times and momentarily identify with Mendel, perhaps they can grasp the magnitude of resistance and the utter frustration that is involved in attempting to explain something that is novel to someone who is simply incapable of understanding — but someone who theoretically should be capable of understanding. Neither Nageli nor anyone else had any comprehension of what Mendel had discovered, no matter how Mendel explained himself, until quite independently three individuals rediscovered genetics. Others had to repeat Mendel's work, that is, *they had to experience what he had done*, before they could understand what he had said. All scientists should realize that when they review others' work, they are potential Nagelis. But characteristically scientists persist in failing to learn from history and so failing are condemned to repeat the past. Surely the "expert" must realize that if the history of science teaches nothing more than is contained in the Nageli-Mendel incident, it behooves him to admit to the possibility that he himself may be a little short of omniscient. Perhaps with this first admission, he can go on to teach himself to say, as Galileo taught himself to say, "I do not know".

References

1. Zuckerman, H. and Merton, R., Institutionalized patterns of evaluation in science, in *The Sociology of Science*, ed. R. Merton, University of Chicago Press, Chicago, p.470 (1973).
2. Merton, R., *Social Theory and Social Structure*, The Free Press, New York (1957).

3. Merton, R., *The Sociology of Science*, University of Chicago Press, Chicago (1973). See this and the previous reference for the sociology and sociology of science paradigm as well as for references to other pertinent literature.
4. Newman, J.R., *Science and Sensibility*, Simon, Schuster, New York (1961). Also see ref. 3, pp.267-278.
5. Ref. 1, p.461.
6. Ref. 1, p.495.
7. Darwin, C., *The Autobiography of Charles Darwin and Selected Letters*, Dover, New York (1892).
8. Polanyi, M., The potential theory of adsorption, *Science*, 141, 1010-1012 (1963). Also see refs. 2 and 3.
9. Examples of rejection and resistance are well documented. See: Barber, B., Resistance by scientists to scientific discovery, in *The Sociology of Science*, (B. Barber and W. Hirsch, eds.), The Free Press, Glencoe, pp.539-556 (1962); Dingle, H., *Science at the Crossroads*, Martin Brian and O'Keeffe, London (1972); Stent, G., Prematurity and uniqueness in scientific discovery, *Sci. Am.*, 227, 84-93 (1972); Van Valen, L. and Pitelka, F., Intellectual censorship in ecology, *Ecology*, 55, 925-926 (1974). Also see refs. 1 and 8.
10. Taton, R., *Reason and Chance in Scientific Discovery*, Science Editions, New York, p.147 (1962).
11. See Mulkay, M., Some aspects of cultural growth in the natural sciences, *Social Research*, 36, 22-52 (1969). Our conclusions are very similar to his. See also Lindsey, D., *The Scientific Publication System in Social Science*, Jossey-Bass, San Francisco (1978), for a critical study of some aspects of the referee system in the social sciences. This publication also contains a literature review.
12. Kuhn, T., *The Structure of Scientific Revolutions*, University of Chicago Press (1970).
13. Royce, J., *The Encapsulated Man*, D. Van Nostrand, Princeton (1964).
14. Hoyle, F., *Man in the Universe*, Columbia University Press, New York, p.20 (1966).
15. Kantor, J.R., *Problems of Physiological Psychology*, Principia Press, Chicago, p.93 (1947).
16. Kantor, J.R., *The Scientific Evolution of Psychology*, Vols. I and II, Principia Press, Chicago, (1963 and 1969).
17. Koestler, A., *The Ghost in the Machine*, Macmillan, New York (1967).
18. Ref. 17, p.289.
19. Ref. 17, p.264.
20. Ref. 17, p.179.
21. Kantor, J.R., *The Aim and Progress of Psychology and Other Sciences*, Principia Press, Chicago, p.22 (1971).
22. Kantor, J.R., Cognition as events and as psychic constructs, *Psych. Record*, 28, 329-342 (1978).
23. Van Valen, L. and Pitelka, F., ref. 9.
24. Storer, N.W., *The Social System of Science*, Holt, Rinehart and Winston, pp.119-120 (1966).
25. Ref. 24, p.119.

Dear Mr. Catt

Thank you for your letter and reprint. I'm glad that the referee paper was of some interest. The best work on the subject is Lindsey (see note 11) who summarizes most work up to 1978 or thereabouts. I don't know of any major work on this that has been done since then although one can find dissenting voices usually in the form of letters to editors in such journals as the *New Scientist*: eg. *Physics Today*, 1979 (April), Vol 32, 14-15, *New Scientist*, 16 July 1981, 178, but these amount to very little. I would be interested in your "The rise and fall...." paper.

Michael

P.S. Look particularly at the Mahoney references in the Lindsey book.